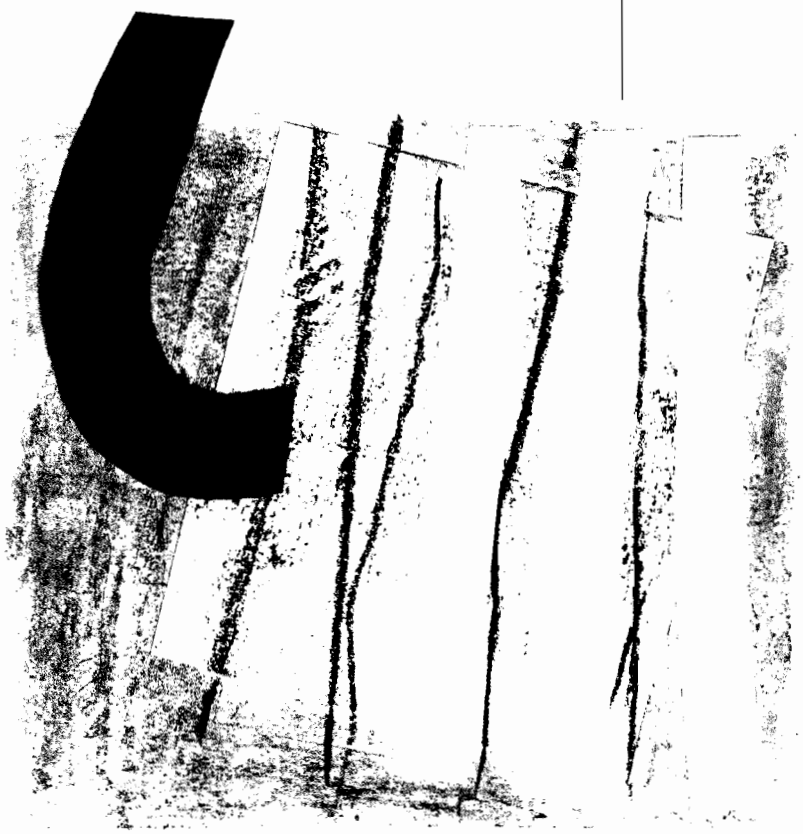
EXPLAINING THE SADDLEPOINT APPROXIMATION

Constantinos Goutis
and George Casella

95-63



WORKING PAPERS

Universidad Carlos III de Madrid

Working Paper 95-63
Statistics and Econometrics Series 27
December 1995

Departamento de Estadística y Econometría
Universidad Carlos III de Madrid
Calle Madrid 126
28903 Getafe (Spain)
Fax (341) 624-9849

EXPLAINING THE SADDLEPOINT APPROXIMATION

Constantinos Goutis and George Casella*

Abstract

Saddlepoint approximations are powerful tools for obtaining accurate expressions for densities and distribution functions. We give an elementary motivation and explanation of saddlepoint approximation techniques, stressing the connection with the familiar Taylor series expansions and the Laplace approximation of integrals. Saddlepoint methods are applied to the convolution of simple densities and, using the Fourier inversion formula, the saddlepoint approximation to the density of a random variable is derived. We then apply the method to densities of sample means of iid random variables, and also demonstrate the technique for approximating the density of a maximum likelihood estimator in exponential families.

Key words and phrases : Laplace method, maximum likelihood estimators, moment generating functions, Taylor series.

* Goutis, Departamento de Estadística y Econometría, Universidad Carlos III de Madrid, tel. (341) 624-9852, e-mail: *costas@est-econ.uc3m.es*; Casella, Biometrics Unit, Cornell University, Ithaca NY 14853, USA. The second author's research was supported by NSF Grant DMS 9305547. We would like to thank Luis Tenorio for useful discussions.

1 Introduction

The saddlepoint approximation has been a valuable tool in asymptotic analysis. Various techniques of accurate approximations, relying on it in one or another way, have been developed since the seminal paper by Daniels (1954). Reid (1988, 1991) gives a comprehensive review of the applications and a broad coverage of the relevant literature.

Typically, derivations and implementations of saddlepoint approximations rely on tools such as exponential tilting, Edgeworth expansions, Hermite polynomials, complex integration and other advanced notions. Although these are important tools for researchers in the area, they tend to obscure the basic underlying premise of the saddlepoint approximation. A goal of this paper is to give an elementary motivation of the technique, stressing the familiar Taylor series expansions. At the beginning we will somewhat ignore the statistical applications, because saddlepoint approximations are general techniques, and quite often reference to random variables and distributions can be more confusing than illuminating. However, once the approximation is developed we will examine some statistical applications.

Throughout the paper, we assume that the functions are as regular as needed. In other words, when we write a derivative or an integral, we assume that they exist. In order to keep the technical level reasonable, we avoid any rigorous asymptotic analysis. Instead, we hint at the quality of the various expansions. Furthermore, we develop the methods in the univariate case. This is almost without loss of generality as the multivariate case is essentially the same but with a somewhat more complicated notation.

In Section 2 we develop the saddlepoint approximation from the point of Taylor series and Laplace approximations, and give some examples. Section 3 is an attempt to explain the original derivation of the saddlepoint, which has its roots in Fourier transforms and complex analysis. Those unwilling to wade through these derivations need only look at formulas (3.10), (3.14) and (3.15), which give the formulas for the density approximations for the sum, and for the MLE in exponential families. Section 4 contains a short discussion.

2 Various Expansions

The saddlepoint approximation arises from a natural sequence of approximations that become progressively more local. In this section we first follow a path that leads to the saddlepoint, then look at some examples.

2.1 From Taylor Series to Saddlepoints

Consider a positive function $f(x)$ and suppose that we would like to approximate its value at some point x_0. Perhaps the simplest way to do it is to use the first few terms of its Taylor series expansion. We will use that idea, but not for $f(x)$ itself but for $h(x) \equiv \log f(x)$. Writing $f(x) = \exp h(x)$ and choosing x_0 as the point to expand around, we obtain

$$f(x) \approx \exp\left\{h(x_0) + (x - x_0)h'(x_0) + \frac{(x - x_0)^2}{2}h''(x_0)\right\}. \tag{2.1}$$

The above approximation simplifies if we choose $x_0 = \hat{x}$, where $h'(\hat{x}) = 0$. The second term disappears and we have

$$f(x) \approx \exp\left\{h(\hat{x}) + \frac{(x - \hat{x})^2}{2}h''(\hat{x})\right\}. \tag{2.2}$$

Equality (2.2) will be exact if $h(x)$ is a quadratic function. If not, obviously for an x far from $\hat{x}$, the omitted terms of order $(x - \hat{x})^3$ and higher will be important and the approximation will not be good.

The approximation (2.2) can be useful in its own right, but it can also be used for computing integrals of positive functions, such as $\int f(x)dx$. Expanding the integrand as in (2.2), we obtain

$$\int f(x)dx \approx \int \exp\left\{h(\hat{x}) + \frac{(x - \hat{x})^2}{2}h''(\hat{x})\right\} dx. \tag{2.3}$$

If $\hat{x}$ is a maximum, $h''(\hat{x})$ is negative and the right side of (2.3) can be explicitly computed by recognizing that the kernel of the integral is the same as the kernel of a normal density with mean $\hat{x}$ and variance $-1/h''(\hat{x})$. Hence

$$\int f(x)\,dx \approx \exp\{h(\hat{x})\}\left(-\frac{2\pi}{h''(\hat{x})}\right)^{1/2}. \tag{2.4}$$

The above technique is called the *Laplace approximation*, and as its name suggests, it has been known for ages. It implicitly requires that the integral is computed over the whole line, but it may be sufficiently accurate as long as the mass of the approximating function is within the limits of the integration.

The next, perhaps most natural step is to combine the two methods (2.2) and (2.4), that is, to try to approximate a function by a Laplace type approximation of an integral. We first write the function f as

$$f(x) = \int m(x,t)\,dt, \tag{2.5}$$

for some positive $m(x,t)$. This is always possible, by considering for example $m(x,t) = f(x)m_0(t)$, where $m_0(t)$ is a function integrating to one, but the latter representation is not particularly useful or illuminating. (We will later see a number of cases where the representation (2.5) arises fairly naturally.)

By defining $k(x,t) = \log m(x,t)$ we consider the Laplace approximation of the integral of $\exp k(x,t)$ with respect to the second variable t. For any fixed x, we write

$$f(x) \approx \int \exp\left\{k(x,\hat{t}(x)) + \frac{(t-\hat{t}(x))^2}{2}\left.\frac{\partial^2 k(x,t)}{\partial t^2}\right|_{\hat{t}(x)}\right\} dt \tag{2.6}$$

$$= \exp\left\{k(x,\hat{t}(x))\right\}\left(-\frac{2\pi}{\left.\frac{\partial^2 k(x,t)}{\partial t^2}\right|_{\hat{t}(x)}}\right)^{1/2} \tag{2.7}$$

where, for each x, $\hat{t}(x)$ satisfies $\partial k(x,t)/\partial t = 0$ and $\partial^2 k(x,t)/\partial t^2 < 0$, and hence maximizes $k(x,t)$. Equation (2.7) is a *saddlepoint approximation* of $f(x)$, and $\hat{t}(x)$ is the *saddlepoint*.

In the above expressions, the notation makes it explicit that the maximum $\hat{t}$ depends on x. In that sense, (2.7) is a set of integrated Taylor expansions, one for each x, as opposed to (2.2) which is a single series around $\hat{x}$. One can hope that, in general, this fact should make the approximation (2.7) more accurate than (2.2). However, there is a price to be paid: if we want the values of $f(x)$ at various points x, we must compute $\hat{t}(x)$, $k(x,\hat{t}(x))$ and $\partial^2 k(x,t)/\partial t^2$ each time afresh.

It is clear that (2.6) will be exact if $k(x,t)$ is a quadratic function of t for each x, but, as we will see, there are other cases where the approximation is

also exact. However, the choice of the form of $k(x,t)$ is usually a function of other considerations. Furthermore, it is worth noting that the accuracy of (2.7) depends on the value of x where we want to approximate $f(x)$ because, in general, the omitted third and higher order derivatives with respect to t depend on x.

2.2 Convolutions

We illustrate the above approximations by deriving the distribution of the sum of two random variables by approximating the convolution integral.

Example: Gamma Distribution. We first look at a simple, somewhat artificial example. Suppose that the function we wish to approximate is a gamma density with shape parameter equal to 2α, (with $\alpha > 1$) and scale parameter equal to 1, that is,

$$f(x) = \frac{1}{\Gamma(2\alpha)} x^{2\alpha-1} \exp(-x) \tag{2.8}$$

for positive x. The Taylor series approximation is obtained by applying (2.2). The maximum is achieved at $\hat{x} = 2\alpha - 1$ and the second derivative of the logarithm of (2.8) evaluated at $\hat{x}$ is equal to $-1/(2\alpha - 1)$. Hence

$$f(x) \approx \frac{1}{\Gamma(2\alpha)} \exp\left\{(2\alpha-1)\left[\log(2\alpha-1) - 1\right] - \frac{(x-2\alpha+1)^2}{2(2\alpha-1)}\right\} \tag{2.9}$$

Recalling that $f(x)$ is actually a density function, we can *renormalize* the approximation by calculating the constant of the right side of (2.9) so that $\int f(x)\,dx = 1$. Doing this, we obtain

$$f(x) \approx \frac{1}{[2\pi(2\alpha-1)]^{1/2}} \exp\left\{-\frac{(x-2\alpha+1)^2}{2(2\alpha-1)}\right\}, \tag{2.10}$$

a normal density with mean and variance both equal to $2\alpha - 1$.

We can probably write $f(x)$ in the form (2.5) in several ways, but a simple one is motivated from elementary distribution theory. We know that the sum of two independent $\Gamma(\alpha, 1)$ random variables is another $\Gamma(2\alpha, 1)$

random variable, so if $g(\cdot)$ is the $\Gamma(\alpha,\ 1)$ density then $f(x)$ is a convolution of the form

$$\begin{aligned} f(x) &= \int_0^x g(x-y)g(y)\,dy, \\ &= \frac{\exp(-x)}{\Gamma(\alpha)^2}\int_0^x (x-y)^{\alpha-1}y^{\alpha-1}\,dy \\ &= \frac{\exp(-x)}{\Gamma(\alpha)^2}\int_0^x \exp\left\{(\alpha-1)\left[\log(x-y)+\log y\right]\right\}\,dy\,. \end{aligned} \tag{2.11}$$

The exponent in the integrand is maximized for $y = x/2$ and its second derivative with respect to y is equal to $-(\alpha-1)\left\{(x-y)^{-2}+y^{-2}\right\}$. Applying (2.7) we obtain

$$\int \exp\left\{(\alpha-1)\left[\log(x-y)+\log y\right]\right\}\,dy \approx \left(\frac{\pi}{\alpha-1}\right)^{1/2}\frac{x^{2\alpha-1}}{2^{2\alpha}}. \tag{2.12}$$

Substituting in (2.11), we see that the kernel of the approximating function is a gamma density with a shape parameter equal to $2\alpha-1$ and scale parameter equal to one. Thus, if we renormalize it so that it integrates to one we obtain the $\Gamma(2\alpha,\ 1)$ density exactly. This is one of the three cases where the saddlepoint approximation (to the density of a sample mean or sum) is exact. The other two cases are the normal and the inverse Gaussian distributions (see Daniels 1980).

It is interesting to note that the convolution formula (2.11) can be solved for the saddleppoint in some generality. If we write

$$k(x,y) = \log[g(x-y)g(y)] = \log[g(x-y)] + \log[g(y)],$$

then under mild regularity conditions, $\partial k(x,y)/\partial y$ has a zero at $y = x/2$, and we can apply the convolution formula somewhat straightforwardly to other densities.

Example: Student's t distribution. Let X be a random variable with a Student's t distribution with ν degrees of freedom ($X \sim t_\nu$). Then X has density

$$f(x) = C_\nu \frac{1}{(1+x^2/\nu)^{\frac{\nu+1}{2}}}\,, \text{ where } C_\nu = \frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\Gamma\left(\frac{\nu}{2}\right)\sqrt{\nu\pi}}\,.$$

To obtain the density of the sum $T = X_1 + X_2$, where X_1 and X_2 are independent t_ν random variables, we use (2.11) and find the zeros of

$$k(x,y) = \log\left(1 + (x-y)^2/\nu\right) + \log\left(1 + y^2/\nu\right).$$

The three solutions of $\partial k(x,y)/\partial y = 0$ are $y = x/2$ and $y = (x \pm \sqrt{x^2 - 4\nu})/2$. For $x^2 < 4\nu$ the last two roots are complex, and $\hat{y} = x/2$ is the saddlepoint. Applying (2.7) we obtain the saddlepoint approximation to the distribution of T as

$$f_T(x) \approx C_\nu^2 \sqrt{2\pi}\nu \left[\frac{4\nu}{x^2 + 4\nu}\right]^\nu \left[\frac{1}{4\nu - x^2}\right]^{1/2}. \tag{2.13}$$

To evaluate the approximation, we compare it to the *exact* density of T for $\nu = 9$[1], which is

$$f_T(x) = \frac{C_\nu^2 5\pi\sqrt{9}}{64} \frac{622336 + 56576\frac{x^2}{9} + 48496\frac{x^4}{9^2} + 272\frac{x^6}{9^3} + 7\frac{x^8}{9^4}}{(x^2/9 + 4)^9}. \tag{2.14}$$

Figure 2.2 compares the saddlepoint approximation, and the renormalized saddlepoint approximation, with the exact density. The agreement is quite good.

2.3 Approximating Marginals

Another application of the saddlepoint approximation arises quite naturally in the calculation of marginal densities. For example, if $(X,Y) \sim f(x,y)$, then by a direct application of (2.7) the marginal density of X, $f_X(x)$ can be approximated by

$$\begin{aligned} f_X(x) &= \int f(x,y)\,dy \\ &\approx \sqrt{2\pi} f(x,\hat{y}) \left(\left. \frac{\partial^2 \log f(x,y)}{\partial y^2} \right|_{\hat{y}} \right)^{-1/2} \end{aligned} \tag{2.15}$$

where $\hat{y} = \hat{y}(x)$ satisfies $\partial \log f(x,y)/\partial y = 0$ and is a local maximum. It is easy to see that by repeating the saddlepoint approximation, we can also marginalize higher dimensional densities.

[1]The exact density of the sum of two iid t_ν random variables was found with Roger Berger during a fit of *Mathematica*-induced frenzy.

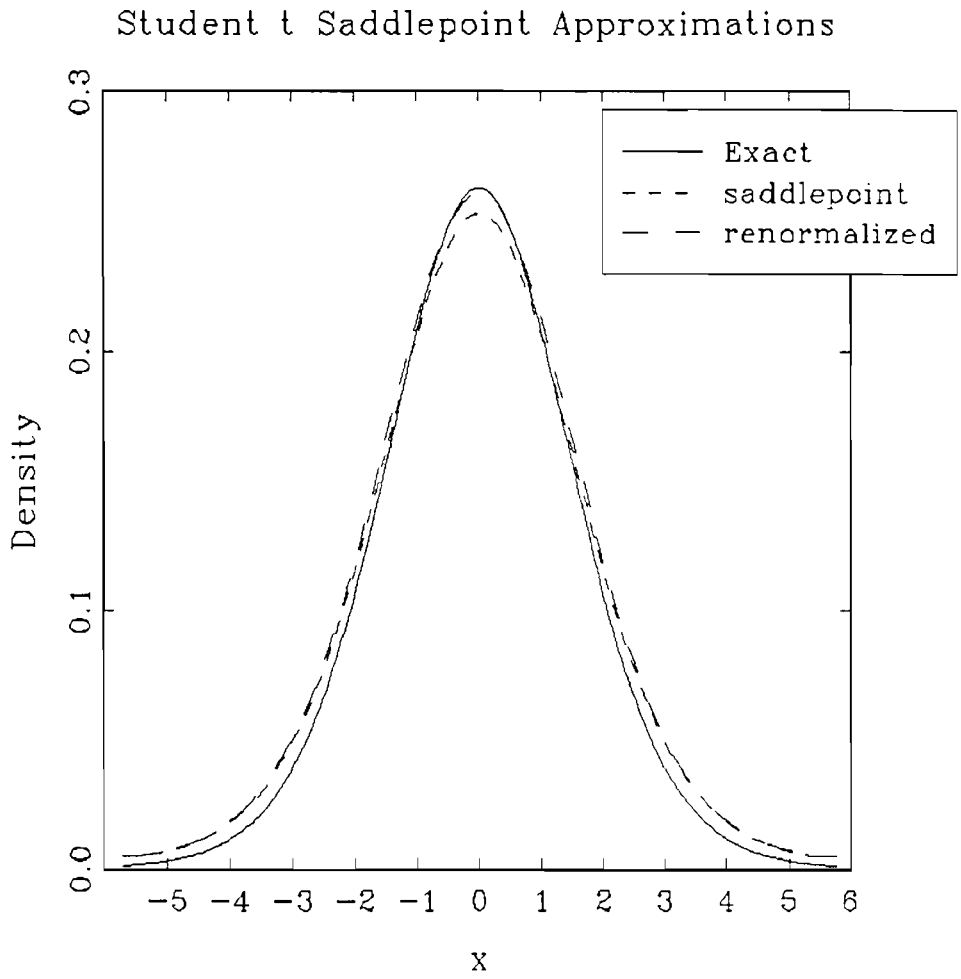


Figure 2.1: Exact density of a t_9 (solid line) together with the saddlepoint approximation (short dashes) and renormalized saddlepoint approximation (long dashes). The two saddlepoint approximations are virtually the same with the exception of the middle region, where the renormalized one is better.

Formula (2.15) can also be applied to marginal likelihoods, a desirable enterprise in the presence of nuisance parameters. Given a likelihood function $L(\theta, \lambda|\mathbf{x})$, where θ is the parameter of interest, λ is a nuisance parameter, and $\mathbf{x}$ is the data, a marginal likelihood L_M for θ is obtained by integrating out λ. This again puts us in the situation of a saddlepoint approximation and, similar to (2.15) we can write

$$\begin{aligned} L_M(\theta|\mathbf{x}) &= \int L(\theta, \lambda|\mathbf{x})\, d\lambda \\ &\approx L(\theta, \hat{\lambda}|\mathbf{x}) \left(\frac{\partial^2 \log L(\theta, \lambda|\mathbf{x})}{\partial \lambda^2}\bigg|_{\hat{\lambda}} \right)^{-1/2}. \end{aligned} \tag{2.16}$$

This approximation is the Cox and Reid (1987) *approximate conditional likelihood*, and can be considered a version of the famous *modified profile likelihood* of Barndorff-Nielsen (1983). Note that the factor that is "adjusting" the likelihood is the observed Fisher information. See Barndorff-Nielsen (1988) for an alternate derivation and further discussion.

3 The Real Thing

The examples in the previous section are, of course, rather artificial. The function $f(x)$ has a closed form and there is no need to use any approximations. Indeed, we do not gain anything as the approximating functions are equally complicated as the original one. As one might expect, this is not always the case. In this section we look at a number of important statistical problems where saddlepoint approximations are useful.

The first statistical application of the saddlepoint approximation was derived by Daniels (1954). He approached the problem as the inversion of a Fourier transform. Such an approach has the advantage of automatically providing a function $m(\cdot)$ satisfying (2.5), but also carries the disadvantage of making us deal with complex integration. We begin with some details about the *inversion formula.*

3.1 The Inversion Formula

We recall that for a density $f(x)$, the *moment generating function* is defined as

$$\phi_X(t) = \int_{-\infty}^{+\infty} \exp(tx) f(x)\, dx. \tag{3.1}$$

¿From $\phi_X(t)$, we can obtain $f(x)$ by using the inversion formula (Feller 1971, Chapter XV; Billingsley 1995, Section 26)

$$\begin{aligned} f(x) &= \frac{1}{2\pi} \int_{-\infty}^{+\infty} \phi_X(it) \exp(-itx)\, dt \\ &= \frac{1}{2\pi} \int_{-\infty}^{+\infty} \exp\left\{\kappa_X(it) - itx\right\}\, dt \end{aligned} \tag{3.2}$$

where $i = \sqrt{-1}$ and we have defined $\kappa_X(t) = \log \phi_X(t)$. These formulae are common in statistical contexts where $f(x)$ is a density and $\phi_X(it)$ is the *characteristic function.* (Waller et al. 1995 discuss exact numerical inversion of (3.2).) The function $\kappa_X(t) = \log \phi_X(t)$ is also called the *cumulant generating function* of a random variable X. However, we do not have to think of a random variable at all, and (3.2) is applicable even if $f(x)$ is negative or does not integrate to one. There is a complication in that we have to deal with complex rather than real numbers, but (3.2) is similar to (2.5) and this suggests that we can use the same ideas as in the previous section.

We first make a change of variable ($t' = it$) to obtain

$$f(x) = \frac{1}{2\pi i} \int_{\tau - i\infty}^{\tau + i\infty} \exp\left\{\kappa_X(t) - tx\right\}\, dt. \tag{3.3}$$

for $\tau = 0$. We will need to integrate (3.3) for some $\tau \neq 0$, but, it turns out that the integral (3.3) is the same for any τ near zero where $\phi_X(t)$ exists.

We take $k(x,t) = \kappa_X(t) - tx$ and, as in (2.7), we find the saddlepoint, that is, the point $\hat{t}(x)$ that satisfies

$$\kappa'_X(t) = x. \tag{3.4}$$

Expanding the exponent in (3.3) around $\hat{t}(x)$ we have

$$\kappa_X(t) - tx \approx \kappa_X(\hat{t}(x)) - \hat{t}(x)x - \frac{\left(t - \hat{t}(x)\right)^2}{2} \kappa''_X(\hat{t}(x)). \tag{3.5}$$

We now substitute in (3.3) and integrate with respect to t along the line parallel to the imaginary axis through the point $\hat{t}(x)$, that is, we choose the point τ in the limits of the integral to be $\hat{t}(x)$. Applying expression (2.7) yields

$$f_X(x) \approx \left(\frac{1}{2\pi\, \kappa_X''(\hat{t}(x))}\right)^{1/2} \exp\left\{\kappa_X\left(\hat{t}(x)\right) - \hat{t}(x)x\right\}. \tag{3.6}$$

Note that to do the integration here we require $\kappa_X''(\hat{t}(x)) > 0$ (compare to (2.3), where we needed $\hat{x}$ to be a maximum). Although this inequality is often satisfied, it is not known if it holds in general for a cumulant generating function. Indeed, it is not obvious if (3.4) has a real solution. (Daniels 1954 established that if κ_X is the cumulant generating function of the sample mean of iid random variables, $\kappa_X''(\hat{t}(x))$ will be greater than 0.)

Expression (3.6) is what is commonly thought of as the saddlepoint approximation to a density. Its error of approximation is much better than the Taylor series approximation to a function. (In classical, or "first-order" asymptotics, the error terms usually decrease at the rate $n^{1/2}$, for a sample of size n. The saddlepoint is "second-order" asymptotics, and can have error terms decrease as fast as $n^{3/2}$, which yields a big impro vement in accuracy for small samples.) For more details see the review articles by Reid (1988, 1991), or the books by Kolassa (1994) or McCulloch (1987).

It is important to note that the assumption of the existence of the moment generating function is stronger than needed. It is only necessary that the integral exists for t in a neighborhood of zero, so that (3.1) can be replaced by

$$\phi_X(t) = \int_{-\infty}^{\infty} \exp(tx) f(x)\, dt < \infty \text{ for } c_1 < t < c_2$$

where $c_1 \leq 0$ and $c_2 \geq 0$ and $c_2 - c_1 > 0$. Thus one, but not both, c_i can even be zero. These are the technical conditions that are needed in order to do the complex integration. Nevertheless, the end result is that the saddlepoint approximation can be used on a wide variety of densities, even those whose moment generating function do not exist.

Example: Noncentral chi-squared. An interesting application of the saddlepoint approximation occurs in the case of the noncentral chi-squared

density. This density has no closed form, and is usually written

$$f(x|\lambda) = \sum_{k=0}^{\infty} \frac{x^{p/2+k-1}e^{-x/2}}{\Gamma(p/2+k)2^{p/2+k}} \frac{\lambda^k e^{-\lambda}}{k!}, \tag{3.7}$$

where p is the degrees of freedom and λ is the noncentrality parameter. The density is an infinite mixture of central chi-squared densities, where the weight are Poisson probabilities. It turns out that calculation of the moment generating function is simple, and it can be expressed in closed form as

$$\phi_X(t) = \frac{e^{2\lambda t/(1-2t)}}{(1-2t)^{p/2}}. \tag{3.8}$$

Solving the saddlepoint equation $\partial \log \phi_X(t)/\partial t = x$ yields the saddlepoint

$$\hat{t}(x) = \frac{-p + 2x - \sqrt{p^2 + 8\lambda x}}{4x} \tag{3.9}$$

and applying (3.6) yields the approximate density. For $p = 7$ and $\lambda = 5$ Figure 3.1 displays the saddlepoint, renormalized saddlepoint, and exact densities. Here, the saddlepoint and the renormalized saddlepoint are remarkably accurate. (See Hougaard 1988 for a generalization.)

3.2 Saddlepoints for Sums

The real power of the saddlepoint comes to light when it is applied to the distribution of sums or averages. Perhaps the simplest non-trivial example, which is also the oldest one (Daniels 1954), is the derivation of the density of a sample mean of independent and identically distributed random variable. The key here is that the moment generating function of a sum of iid random variables can be easily computed from the original moment generating function, and (3.6) can be directly applied.

Consider $\bar{X}$ to be the sample mean of $X_1, X_2, \ldots, X_n$, iid random variables. Each X_i has a moment generating function $\phi_X(t)$ and a cumulant generating function $\kappa_X(t)$. An elementary statistical argument shows that the moment generating function of $\bar{X}$ is $\phi_{\bar{X}}(t) = \phi_X(t/n)^n$ and the cumulant generating function is $\kappa_{\bar{X}}(t) = n\kappa_X(t/n)$. A direct application of (3.6) then gives

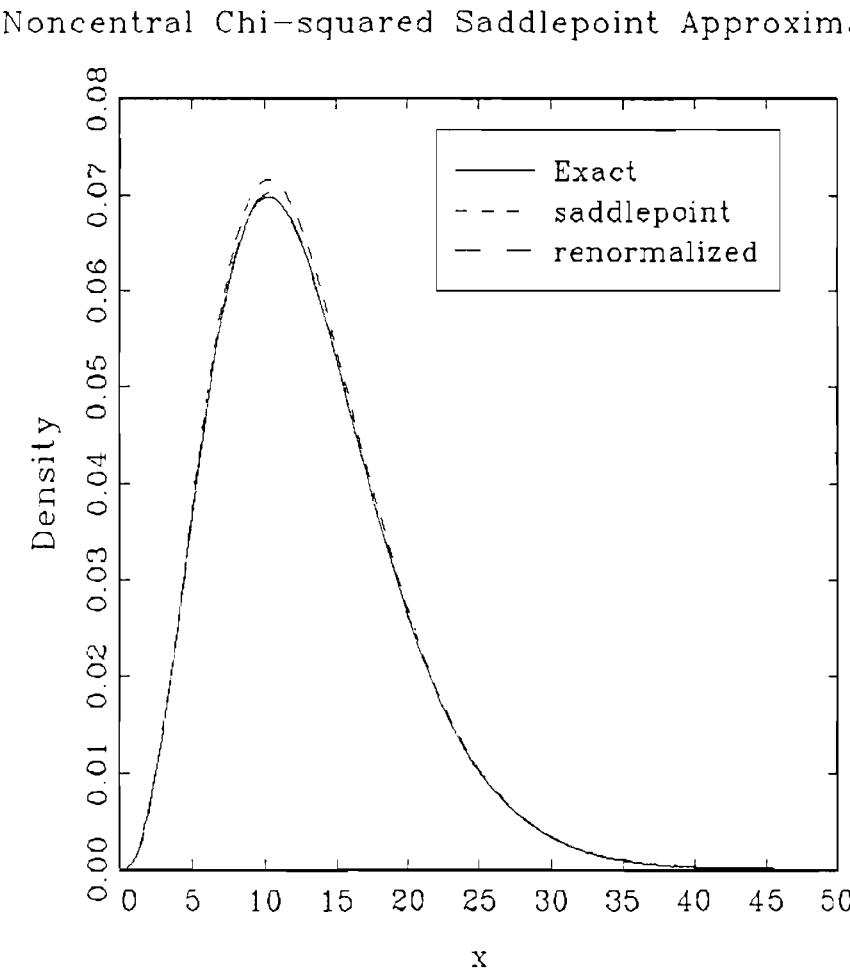


Figure 3.1: Exact density of a χ^2_7 (solid line) together with the saddlepoint approximation (short dashes) and renormalized saddlepoint approximation (long dashes). The renormalized saddlepoint approximations is virtually the same as the exact density.

$$(3.10) \qquad f_{\bar{X}}(\bar{x}) \approx \left(\frac{n}{2\pi\,\kappa_X''(\hat{t}(\bar{x}))}\right)^{1/2} \exp\left\{n\left[\kappa_X\left(\hat{t}(\bar{x})\right) - \hat{t}(\bar{x})\bar{x}\right]\right\}.$$

The right side of expression (3.10) is the famous saddlepoint approximation to the density of $\bar{X}$. Of course, there are several loose ends in the derivation which should be formalized in order to be legitimate. Daniels (1954, 1987) presents all the details. Note that the fact that we are dealing with densities and random variables enters only in the derivation of the cumulant $\kappa_{\bar{X}}(t)$ from the cumulants of the individual random variables. We can also appeal to it to renormalize $f_{\bar{X}}(\bar{x})$ so that it integrates to 1, which amounts to adjusting the constant $(n/(2\pi))^{1/2}$. This usually requires numerical integration.

The saddlepoint approximation can also be used on discrete distributions, as the next example shows.

Example: Poisson Distribution. Let $X_1, \ldots, X_n$ be iid from the Poisson distribution with mean λ. The cumulant generating function of X_i is $\kappa_X(t) = \lambda\,(\exp(t) - 1)$ which yields $\hat{t}(\bar{x}) = \log(\bar{x}/\lambda)$ as the saddlepoint. The formula (3.10) can be used directly, but now the mean can take only values $\bar{x} = r/n$ for integer r. Substituting we obtain

$$\begin{aligned} f_{\bar{X}}(\bar{x}) &\approx \left(\frac{1}{2\pi n\,\bar{x}}\right)^{1/2} \exp\left\{n\left[\lambda\left(\frac{\bar{x}}{\lambda} - 1\right) - \left(\log\frac{\bar{x}}{\lambda}\right)\bar{x}\right]\right\} \\ &= \left(\frac{1}{2\pi n}\right)^{1/2} e^{-\lambda n}\frac{\lambda^{n\bar{x}}}{\bar{x}^{n\bar{x}+1/2}}. \end{aligned}$$

This amounts to replacing the factorial in the exact distribution of $\bar{x}$ by Stirling's approximation.

It is interesting to compare the approximation (3.10) with the simpler one, obtained by expanding the integrand in the inversion formula around 0 rather than around its maximum. If we keep the first two terms we obtain

$$\begin{aligned} f_{\bar{X}}(\bar{x}) &= \frac{1}{2\pi}\int_{-\infty}^{+\infty} \exp\left\{n\kappa_X(it) - it\bar{x}\right\} dt \\ &\approx \frac{1}{2\pi}\int_{-\infty}^{+\infty} \exp\left\{n\left[\kappa_X(0) + (\kappa_X'(0) - \bar{x})\frac{it}{n} - \kappa_X''(0)\frac{t^2}{2n^2}\right]\right\} dt. \end{aligned}$$

This corresponds to the inversion of a quadratic cumulant function $i\kappa'_X(0)t - \kappa''_X(0)t^2/(2n)$, and it can be easily seen that we have nothing more or less than the Central Limit Theorem. So the gain in the accuracy of the saddlepoint approximation over the Central Limit Theorem can be attributed to a more clever expansion of the logarithm of the integrand. This expansion is around a value which depends on $\bar{x}$, the value at which we would like to approximate the density, rather than around 0.

3.3 Maximum Likelihood Estimation

The saddlepoint approximation is extremely useful in obtaining approximations to the density of the maximum likelihood estimator, particularly in exponential families (see Daniels 1983 for generalizations). Consider $X_1, X_2, \ldots, X_n$, independent random variables with density

$$f(x|\theta) = \exp\{\theta s(x) - \kappa(\theta) - d(x)\}. \tag{3.11}$$

A version of the sufficient statistic is $S = \sum_{i=1}^{n} s(X_i)$ which has a density

$$f(s|\theta) = \exp\{\theta s - n\kappa(\theta) - h(s)\}. \tag{3.12}$$

¿From sufficiency considerations, we need to consider only $f(s|\theta)$ but, we now proceed in a slightly different manner.

If we apply the saddlepoint approximation to the entire density (3.12), the saddlepoints will be functions of θ. While this is not inherently bad, it brings in an additional complication especially since the part of (3.12) that involves θ is relatively simple anyway. Thus, we will leave that part alone and apply the saddlepoint approximation only to the function $\exp\{-h(s)\}$.

The first step is to find the Laplace transform and the cumulant generating function of $\exp\{-h(s)\}$. Because $f(s|\theta)$ is a density, it integrates to one, and hence by integrating and rearranging (3.12) it follows that

$$\exp\{n\kappa(\theta)\} = \int \exp(\theta s) \exp\{-h(s)\}\, ds. \tag{3.13}$$

The right side is exactly (3.1) with θ instead of t and s instead of x. Here, we have to think of θ as a dummy variable rather than a parameter of a distribution. Hence, the cumulant of $\exp\{-h(s)\}$ is $n\kappa(\theta)$. Using the saddlepoint

approximation for $\exp\{-h(s)\}$ only, we obtain that the density of $f(s|\theta)$ is approximated by

$$f(s|\theta) \approx \exp\{\theta s - n\kappa(\theta)\} \frac{1}{\left[2\pi n\, \kappa''(\hat{t}(s))\right]^{1/2}} \exp\left\{n\kappa\left(\hat{t}(s)\right) - \hat{t}(s)s\right\}$$

$$= \frac{1}{\left[2\pi n\, \kappa''\left(\hat{t}(s)\right)\right]^{1/2}} \exp\left\{\left[\theta - \hat{t}(s)\right] s - n\left[\kappa(\theta) - \kappa\left(\hat{t}(s)\right)\right]\right\} \tag{3.14}$$

where $\hat{t}(s)$ solves the equation

$$n\kappa'(t) = s. \tag{3.15}$$

Note the very nice feature of the derivation–we never need to compute $h(s)$. Thus, the approximate density of the sum is obtained almost automatically from $\kappa(\theta)$, the cumulant generating function of an individual observation.

Now, it seems almost an accident that if we take the derivative of the loglikelihood corresponding to the density of the sufficient statistic s and set it equal to zero, we obtain (3.15), and, hence, $\hat{t}(s)$ is the maximum likelihood estimate. Equation (3.15) suggests that we can obtain the density of the maximum likelihood estimate by a transformation of $f(s|\theta)$. Since $\hat{t}$ and s are related by $s = n\kappa'(\hat{t})$, the Jacobian of the transformation is $n\kappa''(\hat{t})$. Writing s as $s(\hat{t})$, a function of $\hat{t}$, we obtain

$$f(\hat{t}|\theta) \approx \left[\frac{n\kappa''(\hat{t})}{2\pi}\right]^{1/2} \exp\left\{(\theta - \hat{t})s(\hat{t}) - n\left(\kappa(\theta) - \kappa(\hat{t})\right)\right\} \tag{3.16}$$

This is often referred to as the Barndorff-Nielsen p^* - formula (Barndorff-Nielsen 1983) which can be written in a more concise notation. As written in (3.16), it is clear how the random variable $\hat{t}$ and the parameter θ enter the formula. Though the approximation is quite accurate, it typically requires numerical methods to compute $s(\hat{t})$, $\kappa''(\hat{t})$ and the exact normalizing constant. The approach of Barndorff-Neilsen is somewhat different from the saddlepoint approach, and attains similarly accuracy through conditioning on ancillary statistics. See Barndorff-Nielsen and Cox (1994) for a full development.

As it turns out, the renormalized (3.16) is exact for every simple case that we examined, which was essentially the entire exponential family. To

take full advantage of (3.16) requires more complicated settings than we will look at here, so we content ourselves with a simple, exact example.

Example: Pareto Distribution Let $X_1, \ldots, X_n$ be iid from the Pareto distribution with known lower limit α. The density is

$$f(x|\beta) = \frac{\beta\alpha^\beta}{x^{\beta+1}}, \qquad x > \alpha, \tag{3.17}$$

a member of the exponential family. From (3.12) we see that $s = -\sum \log x_i$, $\kappa(\beta) = -\log(\beta\alpha^\beta)$, and the saddlepoint is given by $\hat{t} = \frac{-n}{s+n\log(\alpha)}$. The saddlepoint approximation is straightforward to compute and, from (3.16), the density of $\hat{\beta}$, the maximum likelihood estimator of β, is approximated by

$$f(\hat{\beta}|\beta) \approx \left[\frac{n}{2\pi}\right]^{1/2} \left(\frac{\beta}{\hat{\beta}}\right)^n e^{(1-\beta/\hat{\beta})} \frac{1}{\hat{\beta}}. \tag{3.18}$$

Figure 3.3 shows this approximation and its renormalized version, which is exact.

4 Discussion

A more detail introduction to saddlepoints, including alternate derivations (such as exponential tilting, which avoids complex analysis) can be found in Reid (1988, 1991). In addition, Reid (1988) contains an annotated bibliography.

More recent applications include finite population models (Wang 1993a), bootstrapping and related confidence methods (Wang 1993b, Booth et al 1992, DiCiccio et al. 1992), approximations to distributions functions (Wood et al. 1993), ANOVA and MANOVA (Butler et al. 1992ab, 1993), prior distributions (Eichenhaur-Herrmann and Ickstadt 1993), generalized linear models (Strawderman et al. 1995), exponential linear models (Fraser et al. 1991), multiparameter exponential families (Pierce and Peters 1992) and studentized means (Daniels and Young 1991).

The number of applications of the saddlepoint approximation is quite impressive, as warrants this extremely powerful approximation. The point of this paper is to illustrate that there is a simple basic idea behind this useful technique. Namely, write the quantity one wishes to approximate as

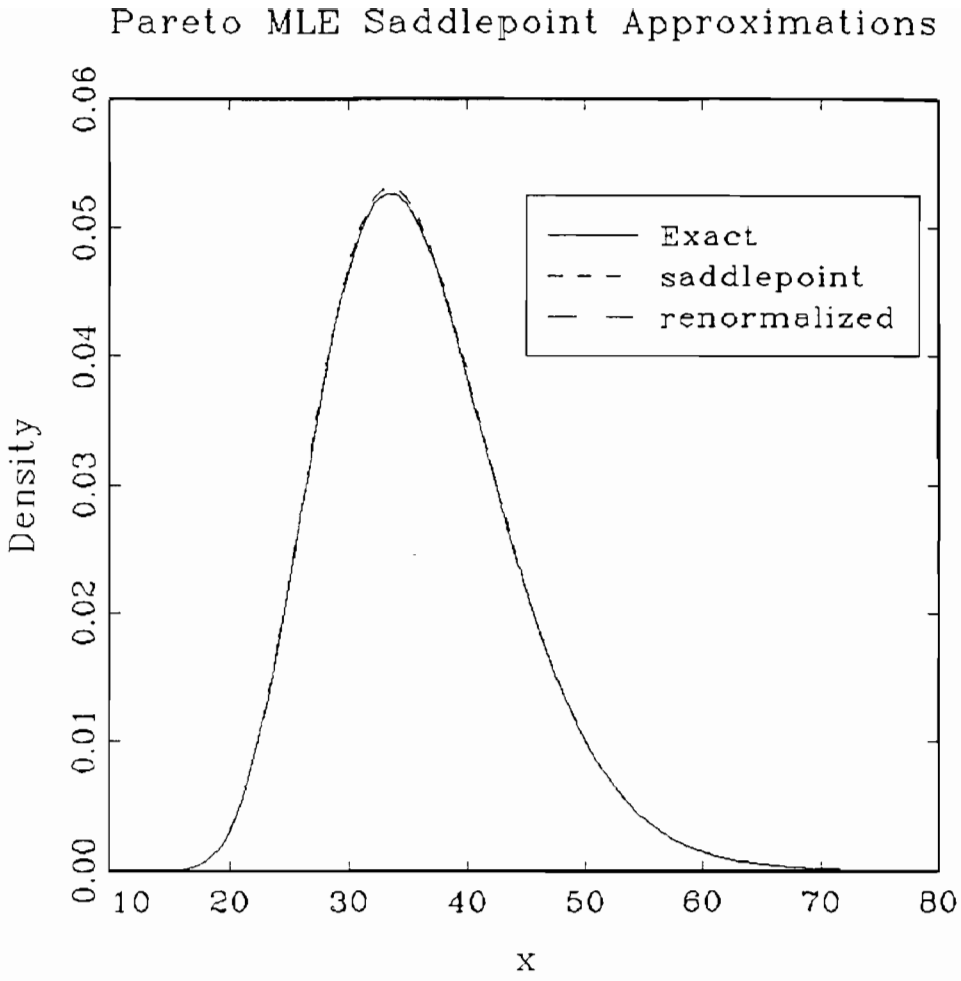


Figure 3.2: Exact density of the MLE from a Pareto distribution based on $n = 10$ observations (solid line) together with the saddlepoint approximation (short dashes) and renormalized saddlepoint approximation (long dashes). The renormalized saddlepoint approximation is exact.

an integral, expand the integrand with respect to the dummy variable of the integral, keep the first few terms and integrate. The integral can be over the complex plane, corresponding to the inversion formula of a Fourier transform but this is a secondary point.

References

[1] Barndorff-Nielsen, O. (1983). On a formula for the distribution of the maximum likelihood estimator. *Biometrika* **70**, 343-365.

[2] Barndorff-Nielsen, O. (1988). Discussion of the paper by Reid. *Statist. Sci.* **3**, 228-229.

[3] Barndorff-Nielsen, O. and Cox, D. R. (1994). *Inferences and Asymptotics.* London: Chapman and Hall.

[4] Billingsley, P. (1995). *Probability and Measure, Third Edition.* New York: Wiley.

[5] Booth, J. G., Hall, P. and Wood, A. (1992). Bootstrap estimation of conditional distributions. *Ann. Statist.* **20**, 1594-1610.

[6] Butler, R. W., Huzurbazar, S. and Booth, J. G. (1992a). Saddlepoint approximations for the generalized variance and Wilks' statistic. *Biometrika* **79**, 157-169.

[7] Butler, R. W., Huzurbazar, S. and Booth, J. G. (1992b). Saddlepoint approximations for the Bartlett-Nonda-Pillai trace statistic in multivariate analysis. *Biometrika* **79**, 705-515.

[8] Butler, R. W., Huzurbazar, S. and Booth, J. G. (1993). Saddlepoint approximations for tests of block independence, sphericity and equal variances and covariances. *J. Roy. Statist. Soc. Ser. B* **55**, 171-183.

[9] Cox, D. R. and Reid, N. (1987). Parameter orthogonality and approximate conditional inference (with discussion). *J. Roy. Statist. Soc. Ser. B* **49**, 1-39.

[10] Daniels, H. E. (1954). Saddlepoint approximations in statistics. *Ann. Math. Statist.* **25**, 631-650.

[11] Daniels, H. E. (1980). Exact saddlepoint approximations. *Biometrika* **67**, 59-63.

[12] Daniels, H. E. (1983). Saddlepoint approximations for estimating equations. *Biometrika* **70**, 89-96.

[13] Daniels, H. E. (1987). Tail Probability Approximations. *Int. Statist. Rev.* **55**, 37–48.

[14] Daniels, H. E. and Young, G. A. (1991). Saddlepoint approximations for the studentized mean. *Biometrika* **78**, 169-179.

[15] DiCiccio, T. J., Martin, M. A. and Young, G. A. (1992) Fast and accurate double bootstrap confidence intervals. *Biometrika* **79**, 285-295.

[16] Eichenhaur-Herrmann, J. and Ickstadt, K. (1993). A saddlepoint characterization for classes of priors with shape-restricted densities. *Statist. Dec.* **11**, 175-179.

[17] Feller, W. (1971). *An Introduction to Probability Theory and its Applications, Volume II.* New York: Wiley.

[18] Fraser, D. A. S., Reid, N. and Wong, A. (1991). Exponential linear models: A two-pass procedure for saddlepoint approximation. *J. Roy. Statist. Soc. Ser. B* **53**, 483-492.

[19] Hougaard, P. (1988). Discussion of the paper by Reid. *Statist. Sci.* **3**, 230-231.

[20] Kolassa, J. E. (1994). *Series Approximation Methods in Statistics.* New York: Springer-Verlag.

[21] McCulloch, P. (1987). *Tensor Methods in Statistics.* London: Chapman and Hall.

[22] Pierce, D. A. and Peters, D. (1992) Practical use of higher order asymptotics for multiparameter exponential families (with discussion). *J. Roy. Statist. Soc. Ser. B* **54**, 701-737.

[23] Reid, N. (1988). Saddlepoint methods and statistical inference (with discussion). *Statist. Sci.* **3**, 213-238.

[24] Reid, N. (1991). Approximations and asymptotics. *Statist. Theory Model., Essays in Honor of D. R. Cox.* London: Chapman and Hall, 287-334.

[25] Strawderman, R. W., Casella, G. and Wells, M. T. (1995). Practical small sample asymptotics for regression problems. *J. Amer. Statist. Assoc.* To appear.

[26] Waller, L. A., Turnbull, B. W. and Hardin, J. M. (1995). Obtaining distribution functions by numerical inversion of characteristic functions. *Amer. Statist.* **49** (4). To appear.

[27] Wang, S. (1993a). Saddlepoint expansions in finite population problems. *Biometrika* **80**, 583-590.

[28] Wang, S. (1993b). Saddlepoint methods for bootstrap confidence bands in nonparametric regression. *Aust. J. Statist.* **35**, 93-101.

[29] Wood, A. T. A., Booth, J. G. and Butler, R. W. (1993). Saddlepoint approximations to the CDF of some statistics with nonnormal limit distributions. *J. Amer. Statist. Assoc.* **88**, 680-686.